

OIEP

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/841,553

DATE: 06/14/2001

TIME: 12:35:12

Input Set : A:\09841553.txt

Output Set: N:\CRF3\06142001\I841553.raw

ENTERED

SEQUENCE LISTING

3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: TAKAKURA, Hikaru

6 MORISHITA, Mio

7 YAMAMOTO, Katsuhiko

8 MITTA, Masanori

9 ASADA, Kiyozo

10 TSUNASAWA, Susumu

11 KATO, Ikunoshin

13 (ii) TITLE OF INVENTION: ~~HYPER~~ THERMOSTABLE PROTEASE GENES

15 (iii) NUMBER OF SEQUENCES: 42

17 (iv) CORRESPONDENCE ADDRESS:

18 (A) ADDRESSEE: Browdy and Neimark

19 (B) STREET: 419 Seventh Street N.W., Ste. 300

20 (C) CITY: Washington

21 (D) STATE: D.C.

22 (E) COUNTRY: United States of America

23 (F) ZIP: 20004

25 (v) COMPUTER READABLE FORM:

26 (A) MEDIUM TYPE: Floppy disk

27 (B) COMPUTER: IBM PC compatible

28 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

29 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

31 (vi) CURRENT APPLICATION DATA:

C--> 32 (A) APPLICATION NUMBER: US/09/841,553

C--> 33 (B) FILING DATE: 24-Apr-2001

C--> 39 (vii) PRIOR APPLICATION DATA:

36 (A) APPLICATION NUMBER: US 08/894,818

37 (B) FILING DATE:

40 (A) APPLICATION NUMBER: JP 323285/1995

41 (B) FILING DATE: 12-DEC-1995

43 (viii) ATTORNEY/AGENT INFORMATION:

44 (A) NAME: Browdy, Roger L.

45 (B) REGISTRATION NUMBER: 25,618

46 (C) REFERENCE/DOCKET NUMBER: TAKAKURA=1

48 (ix) TELECOMMUNICATION INFORMATION:

49 (A) TELEPHONE: (202) 628-5197

50 (B) TELEFAX: (202) 737-3528

53 (2) INFORMATION FOR SEQ ID NO: 1:

55 (i) SEQUENCE CHARACTERISTICS:

56 (A) LENGTH: 659 amino acids

57 (B) TYPE: amino acid

58 (C) STRANDEDNESS: single

59 (D) TOPOLOGY: linear

61 (ii) MOLECULE TYPE: peptide

63 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

65 Met Lys Arg Leu Gly Ala Val Val Leu Ala Leu Val Leu Val Gly

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/841,553

DATE: 06/14/2001

TIME: 12:35:12

Input Set : A:\09841553.txt

Output Set: N:\CRF3\06142001\I841553.raw

66		5		10		15
68	Leu	Leu	Ala	Gly	Thr	Ala
69		20		25		30
71	Arg	Asn	Asn	Ala	Val	Gln
72		35		40		45
74	Gly	Leu	Phe	Lys	Lys	Val
75		50		55		60
77	Asp	Thr	Val	Ile	Met	Phe
78		65		70		75
80	Val	Lys	Val	Leu	Arg	Leu
81		80		85		90
83	Lys	Ile	Ile	Pro	Ala	Val
84		95		100		105
86	Leu	Leu	Ile	Ala	Gly	Met
87		110		115		120
89	Arg	Val	Ser	Gly	Ile	Lys
90		125		130		135
92	Val	Asp	Asp	Ala	Thr	Ser
93		140		145		150
95	Trp	Asn	Ser	Leu	Gly	Tyr
96		155		160		165
98	Val	Asp	Thr	Gly	Ile	Asp
99		170		175		180
101	Val	Ile	Gly	Trp	Tyr	Asp
102		185		190		195
104	Asp	Asp	Gln	Gly	His	Gly
105		200		205		210
107	Thr	Gly	Ser	Val	Asn	Ser
108		215		220		225
110	Lys	Leu	Val	Gly	Val	Lys
111		230		235		240
113	Val	Ser	Thr	Ile	Ile	Ala
114		245		250		255
116	Asp	Lys	Tyr	Gly	Ile	Arg
117		260		265		270
119	Gln	Ser	Ser	Asp	Gly	Thr
120		275		280		285
122	Ala	Trp	Asp	Ala	Gly	Ile
123		290		295		300
125	Gly	Pro	Asn	Thr	Tyr	Thr
126		305		310		315
128	Val	Ile	Thr	Val	Gly	Ala
129		320		325		330
131	Phe	Ser	Ser	Arg	Gly	Pro
132		335		340		345
135	Val	Val	Ala	Pro	Gly	Val
136		350		355		360
138	Thr	Ser	Met	Gly	Thr	Pro
139		365		370		375

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/841,553

DATE: 06/14/2001

TIME: 12:35:12

Input Set : A:\09841553.txt

Output Set: N:\CRF3\06142001\I841553.raw

```

141 Gly Thr Ser Met Ala Thr Pro His Val Ser Gly Val Gly Ala Leu
142                               380                               385                               390
144 Ile Leu Gln Ala His Pro Ser Trp Thr Pro Asp Lys Val Lys Thr
145                               395                               400                               405
147 Ala Leu Ile Glu Thr Ala Asp Ile Val Ala Pro Lys Glu Ile Ala
148                               410                               415                               420
150 Asp Ile Ala Tyr Gly Ala Gly Arg Val Asn Val Tyr Lys Ala Ile
151                               425                               430                               435
153 Lys Tyr Asp Asp Tyr Ala Lys Leu Thr Phe Thr Gly Ser Val Ala
154                               440                               445                               450
156 Asp Lys Gly Ser Ala Thr His Thr Phe Asp Val Ser Gly Ala Thr
157                               455                               460                               465
159 Phe Val Thr Ala Thr Leu Tyr Trp Asp Thr Gly Ser Ser Asp Ile
160                               470                               475                               480
162 Asp Leu Tyr Leu Tyr Asp Pro Asn Gly Asn Glu Val Asp Tyr Ser
163                               485                               490                               495
165 Tyr Thr Ala Tyr Tyr Gly Phe Glu Lys Val Gly Tyr Tyr Asn Pro
166                               500                               505                               510
168 Thr Ala Gly Thr Trp Thr Val Lys Val Val Ser Tyr Lys Gly Ala
169                               515                               520                               525
171 Ala Asn Tyr Gln Val Asp Val Val Ser Asp Gly Ser Leu Ser Gln
172                               530                               535                               540
174 Ser Gly Gly Gly Asn Pro Asn Pro Asn Pro Asn Pro Asn Pro Thr
175                               545                               550                               555
177 Pro Thr Thr Asp Thr Gln Thr Phe Thr Gly Ser Val Asn Asp Tyr
178                               560                               565                               570
180 Trp Asp Thr Ser Asp Thr Phe Thr Met Asn Val Asn Ser Gly Ala
181                               575                               580                               585
183 Thr Lys Ile Thr Gly Asp Leu Thr Phe Asp Thr Ser Tyr Asn Asp
184                               590                               595                               600
186 Leu Asp Leu Tyr Leu Tyr Asp Pro Asn Gly Asn Leu Val Asp Arg
187                               605                               610                               615
189 Ser Thr Ser Ser Asn Ser Tyr Glu His Val Glu Tyr Ala Asn Pro
190                               620                               625                               630
192 Ala Pro Gly Thr Trp Thr Phe Leu Val Tyr Ala Tyr Ser Thr Tyr
193                               635                               640                               645
195 Gly Trp Ala Asp Tyr Gln Leu Lys Ala Val Val Tyr Tyr Gly
196                               650                               655

```

199 (2) INFORMATION FOR SEQ ID NO: 2:

201 (i) SEQUENCE CHARACTERISTICS:

202 (A) LENGTH: 1977 base pairs

203 (B) TYPE: nucleic acid

204 (C) STRANDEDNESS: double

205 (D) TOPOLOGY: linear

207 (ii) MOLECULE TYPE: genomic DNA

209 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

211 ATGAAGAGGT TAGGTGCTGT GGTGCTGGCA CTGGTGCTCG TGGGTCTTCT GGCCGGAACG 60
213 GCCCTTGCGG CACCCGTAAA ACCGGTTGTC AGGAACAACG CGGTTTCAGCA GAAGAACTAC 120
215 GGAAGTCTGA CCCCAGGACT GTTCAAGAAA GTCCAGAGGA TGAAGTGGAA CCAGGAAGTG 180

```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/841,553

DATE: 06/14/2001
TIME: 12:35:12

Input Set : A:\09841553.txt
Output Set: N:\CRF3\06142001\I841553.raw

```

217 GACACCGTCA TAATGTTCCG GAGCTACGGA GACAGGGACA GGGCGGTAA GGTACTGAGG 240
219 CTCATGGGCG CCCAGGTCAA GTACTCCTAC AAGATAATCC CTGCTGTCGC GGTAAAAATA 300
221 AAGGCCAGGG ACCTTCTGCT GATCGCGGGC ATGATAGACA CGGGTTACTT CGGTAACACA 360
223 AGGGTCTCGG GCATAAAGTT CATAAGGAG GATTACAAGG TTCAGGTTGA CGACGCCACT 420
225 TCCGTCTCCC AGATAGGGGC CGATACCGTC TGGAAGTCCC TCGGCTACGA CGGAAGCGGT 480
227 GTGGTGGTTG CCATCGTCGA TACGGGTATA GACGCGAACC ACCCCGATCT GAAGGGCAAG 540
229 GTCATAGGCT GGTACGACGC CGTCAACGGC AGGTCGACCC CCTACGATGA CCAGGGACAC 600
231 GGAACCCACG TTGCGGGTAT CGTTGCCGGA ACCGGCAGCG TTAAGTCCCA GTACATAGGC 660
233 GTCGCCCCCG GCGCGAAGCT CGTCGGCGTG GTCGTCCAGA ACAAGGACAA GTACGGGATA 720
235 GTCTCCACCA TCATCGCGGG TGTGACTGG GTCGTCCAGA ACAAGGACAA GTACGGGATA 780
237 AGGGTCATCA ACCTCTCCCT CGGCTCCTCC CAGAGCTCCG ACAGGACCGA CTCCCTCAGT 840
239 CAGGCCGTCA ACAACGCCTG GGACGCCGGT ATAGTAGTCT GCGTCGCCGC CGGCAACAGC 900
241 GGGCCGAACA CCTACACCGT CGGCTCAGCC GCCGCCGCGA GCAAGGTCAT AACCCTCGGT 960
243 GCAGTTGACA GCAACGACAA CATCGCCAGC TTCTCCAGCA GGGGACCGAC CGCGGACGGA 1020
245 AGGCTCAAGC CGGAAGTCGT CGCCCCGCGC GTTGACATCA TAGCCCCGCG CGCCAGCGGA 1080
247 ACCAGCATGG GCACCCCGAT AAACGACTAC TACACCAAGG CCTCTGGAAC CAGCATGGCC 1140
249 ACCCCGCACG TTTCCGGCGT TGGCGCGCTC ATCTCCAGG CCCACCCGAG CTGGACCCCG 1200
251 GACAAGGTGA AGACCGCCCT CATCGAGACC GCCGACATAG TCGCCCCCAA GGAGATAGCG 1260
253 GACATCGCCT ACGGTGCGGG TAGGGTGAAC GTCTACAAGG CCATCAAGTA CGACGACTAC 1320
255 GCCAAGCTCA CCTTCACCGG CTCCGTCGCC GACAAGGGAA GCGCCACCCA CACCTTCGAC 1380
257 GTCAGCGGCG CCACCTTCGT GACCGCCACC CTCTACTGGG ACACGGGCTC GAGCGACATC 1440
259 GACCTCTACC TCTACGACCC CAACGGGAAC GAGGTTGACT ACTCCTACAC CGCCTACTAC 1500
261 GGCTTCGAGA AGGTGCGGTA CTACAACCCG ACCGCCGGAA CCTGGACGGT CAAGGTCGTC 1560
263 AGCTACAAGG GCGCGGCGAA CTACCAAGTC GACGTCGTCA GCGACGGGAG CCTCAGCCAG 1620
265 TCCGGCGGCG GCAACCCGAA TCCAAACCCC AACCCTGACC CAACCCCGAC CACCGACACC 1680
267 CAGACCTTCA CCGGTTCCGT TAACGACTAC TGGGACACCA GCGACACCTT CACCATGAAC 1740
269 GTCAACACCG GTGCCACCAA GATAACCGGT GACCTGACCT TCGATACTTC CTACAACGAC 1800
271 CTCGACCTCT ACCTCTACGA CCCCACCGG AACCTCGTTG ACAGGTCCAC GTCGAGCAAC 1860
273 AGCTACGAGC ACGTCGAGTA CGCCAACCCC GCCCGGGGAA CCTGGACGTT CCTCGTCTAC 1920
275 GCCTACAGCA CCTACGGCTG GGCGGACTAC CAGCTCAAGG CCGTCGTCTA CTACGGG 1977
278 (2) INFORMATION FOR SEQ ID NO: 3:
280 (i) SEQUENCE CHARACTERISTICS:
281 (A) LENGTH: 522 amino acids
282 (B) TYPE: amino acid
283 (C) STRANDEDNESS: single
284 (D) TOPOLOGY: linear
286 (ii) MOLECULE TYPE: peptide
C--> 288 (ix) FEATURE:
W--> 289 OTHER INFORMATION:
W--> 290 Xaa at position 428 is Gly or Val.
291 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
293 Ala Glu Leu Glu Gly Leu Asp Glu Ser Ala Ala Gln Val Met Ala
294 5 10 15
296 Thr Tyr Val Trp Asn Leu Gly Tyr Asp Gly Ser Gly Ile Thr Ile
297 20 25 30
299 Gly Ile Ile Asp Thr Gly Ile Asp Ala Ser His Pro Asp Leu Gln
300 35 40 45
302 Gly Lys Val Ile Gly Trp Val Asp Phe Val Asn Gly Arg Ser Tyr
303 50 55 60

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/841,553

DATE: 06/14/2001

TIME: 12:35:12

Input Set : A:\09841553.txt

Output Set: N:\CRF3\06142001\I841553.raw

```

305 Pro Tyr Asp Asp His Gly His Gly Thr His Val Ala Ser Ile Ala
306                      65                      70                      75
308 Ala Gly Thr Gly Ala Ala Ser Asn Gly Lys Tyr Lys Gly Met Ala
309                      80                      85                      90
311 Pro Gly Ala Lys Leu Ala Gly Ile Lys Val Leu Gly Ala Asp Gly
312                      95                      100                     105
314 Ser Gly Ser Ile Ser Thr Ile Ile Lys Gly Val Glu Trp Ala Val
315                      110                     115                     120
317 Asp Asn Lys Asp Lys Tyr Gly Ile Lys Val Ile Asn Leu Ser Leu
318                      125                     130                     135
320 Gly Ser Ser Gln Ser Ser Asp Gly Thr Asp Ala Leu Ser Gln Ala
321                      140                     145                     150
323 Val Asn Ala Ala Trp Asp Ala Gly Leu Val Val Val Val Ala Ala
324                      155                     160                     165
326 Gly Asn Ser Gly Pro Asn Lys Tyr Thr Ile Gly Ser Pro Ala Ala
327                      170                     175                     180
329 Ala Ser Lys Val Ile Thr Val Gly Ala Val Asp Lys Tyr Asp Val
330                      185                     190                     195
332 Ile Thr Ser Phe Ser Ser Arg Gly Pro Thr Ala Asp Gly Arg Leu
333                      200                     205                     210
336 Lys Pro Glu Val Val Ala Pro Gly Asn Trp Ile Ile Ala Ala Arg
337                      215                     220                     225
339 Ala Ser Gly Thr Ser Met Gly Gln Pro Ile Asn Asp Tyr Tyr Thr
340                      230                     235                     240
342 Ala Ala Pro Gly Thr Ser Met Ala Thr Pro His Val Ala Gly Ile
343                      245                     250                     255
345 Ala Ala Leu Leu Leu Gln Ala His Pro Ser Trp Thr Pro Asp Lys
346                      260                     265                     270
348 Val Lys Thr Ala Leu Ile Glu Thr Ala Asp Ile Val Lys Pro Asp
349                      275                     280                     285
351 Glu Ile Ala Asp Ile Ala Tyr Gly Ala Gly Arg Val Asn Ala Tyr
352                      290                     295                     300
354 Lys Ala Ile Asn Tyr Asp Asn Tyr Ala Lys Leu Val Phe Thr Gly
355                      305                     310                     315
357 Tyr Val Ala Asn Lys Gly Ser Gln Thr His Gln Phe Val Ile Ser
358                      320                     325                     330
360 Gly Ala Ser Phe Val Thr Ala Thr Leu Tyr Trp Asp Asn Ala Asn
361                      335                     340                     345
363 Ser Asp Leu Asp Leu Tyr Leu Tyr Asp Pro Asn Gly Asn Gln Val
364                      350                     355                     360
366 Asp Tyr Ser Tyr Thr Ala Tyr Tyr Gly Phe Glu Lys Val Gly Tyr
367                      365                     370                     375
369 Tyr Asn Pro Thr Asp Gly Thr Trp Thr Ile Lys Val Val Ser Tyr
370                      380                     385                     390
372 Ser Gly Ser Ala Asn Tyr Gln Val Asp Val Val Ser Asp Gly Ser
373                      395                     400                     405
375 Leu Ser Gln Pro Gly Ser Ser Pro Ser Pro Gln Pro Glu Pro Thr
376                      410                     415                     420
W--> 378 Val Asp Ala Lys Thr Phe Gln Xaa Ser Asp His Tyr Tyr Tyr Asp

```

VERIFICATION SUMMARY

DATE: 06/14/2001

PATENT APPLICATION: US/09/841,553

TIME: 12:35:13

Input Set : A:\09841553.txt

Output Set: N:\CRF3\06142001\I841553.raw

L:32 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:33 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:35 M:220 C: Keyword misspelled or invalid format, [(vii) PRIOR APPLICATION DATA:]
L:39 M:220 C: Keyword misspelled or invalid format, [(vii) PRIOR APPLICATION DATA:]
L:288 M:220 C: Keyword misspelled or invalid format, [(ix) FEATURE:]
L:289 M:246 W: Invalid value of Alpha Sequence Header Field, Data=[Xaa at position 428 is Gly or Val.]
L:291 M:246 W: Invalid value of Alpha Sequence Header Field, [FEATURE:], SeqNo=3
L:378 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:411 M:220 C: Keyword misspelled or invalid format, [(ix) FEATURE:]
L:412 M:246 W: Invalid value of Alpha Sequence Header Field, Data=[N at position 1283 is G or T.]
L:414 M:246 W: Invalid value of Alpha Sequence Header Field, [FEATURE:], SeqNo=4